PAPER • **OPEN ACCESS**

Using deep neural networks for predictive modelling of informal settlements in the context of flood risk

To cite this article: Yue Zhu *et al* 2019 *J. Phys.: Conf. Ser.* **1343** 012032

Using deep neural networks for predictive modelling of informal settlements in the context of flood risk

Yue Zhu[1], Christian Geiß[2], Emily So[3]

[1] Department of Architecture, University of Cambridge, United Kingdom
yz591@cam.ac.uk
[2] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, 82234 Weßling, Germany
christian.geiss@dlr.de
[3] Department of Architecture, University of Cambridge, United Kingdom
ekms2@cam.ac.uk

Abstract. Global climate change has substantially increased the risks of cities being adversely affected by natural hazards such as floods. Among the inhabitants of cities at risk, residents dwelling in informal settlements are the most vulnerable group. To identify the future exposure of informal settlements, we adopt a data-driven model from the machine learning domain to anticipate the growth patterns of formal and informal settlements in flood-prone areas. The potential emergence of informal settlements in Shenzhen, China, is predicted by the proposed method. Then, through an analysis of the flood susceptibility of the predicted informal settlement areas, the emerging vulnerability of Shenzhen towards flooding is revealed.

Key words. climate-resilient cities, neural networks, land use prediction, informal settlements, flood susceptibility.

1. Introduction

1.1. Research background

It has been widely recognized that natural hazards, such as floods, storm surges, and landslides, can bring devastating impact to settlement areas, causing substantial casualties and economic loss. Nowadays, due to climate change, the likelihood of a city being hit by natural hazards has largely increased. Moreover, cities are becoming increasingly exposed due to exponential urban growth. Among these growing settlement areas, the emergence of informal settlements has raised significant concerns due to their poor quality of construction (Scovronick, Lloyd, & Kovats, 2015), limited access to urban facilities, and the restricted choice of site selection. Therefore, these informal settlements have become one of the most vulnerable parts of a city. It should also be noted that, as cities continue to expand, the informal settlements in and around cities are growing with them, in this sense, managing these climate-affected risks has become an acute issue. Most studies map hazard susceptibilities of the past or document the current status rather than exploring the developing trend that affects future vulnerability. Since climate-affected disasters tend to have patterns with respect to return periods, the prediction of future urban vulnerability could foreseeably assist disaster preparedness, thus mitigate the risks in the built environment.

1.2. Research goals

Developing efficient tools to identify the emergence of informal settlements can substantially facilitate the monitoring of changes in urban vulnerability. Recent advances in Machine Learning (ML), particularly deep neural networks (DNN), present promising achievements not only in tasks of pattern recognition, but also in predicting trajectories of changes (Schmidhuber, 2015). Such developments in ML provide effective tools for land use classification and make it possible to predict land use changes in the future. The goal of this research is to provide planners and policymakers with methods to assess and mitigate vulnerabilities of urban informality with respect to climate-affected natural hazards.

2. Materials

2.1. Study area

Shenzhen is a coastal city located in the Pearl River delta in southern China. Due to its coastal location and subtropical maritime climate, Shenzhen has long been suffering from climate-affected natural hazards, such as floods and typhoons. According to a World Bank report (2013), Shenzhen has been ranked as one of the 10th most vulnerable cities that are prone to floods. Research also shows that there is a tendency that the frequency and intensity of floods in Shenzhen are increasing (Yang, Scheffran, Qin, & You, 2015). Such phenomenon could be linked to the land use changes over years, including the disappearance of the large amount of mangrove forest and the aggregating of landfills along the coast of Shenzhen (Kimmelman, 2017). The vulnerability of Shenzhen is also constituted in its built environment. Shenzhen was a fishery village which is subject to a rapid urbanization process since 1979. Thereby, the original villages in Shenzhen were merged into urban coverage. Moreover, a large number of migrants contributed to the formation of arrival villages (Taubenböck, Kraff, & Wurm, 2018). As such, the characteristic of the local climate and the vulnerability of the built environment of Shenzhen render it as a suitable case study for this research.

2.2. Data Source

The satellite images used in this research are retrieved from Google Earth, the image resolution is 1meter. The Digital Elevation Map is also collected from Google Earth, with a spatial resolution of 30 meter. The data set contains multiple time steps from the year of 2003 to 2018 at five-year intervals with coverage of Shenzhen and its nearby city Dongguan. In terms of the geographic data, the data of slope aspect and slope gradient is generated based on DEM data and processed in ArcGIS, and the precipitation data is the average annual precipitation from the year 1981 to 2010, which is downloaded from the database of China National Meteorological Information Centre. All the labels in the training data set are manually annotated via an online image annotation tool LabelMe, created by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). The geographic data analysed in this research include elevation data, slope gradients, slope aspects, waterbody distributions and precipitation data.

3. Methods

3.1. Convolutional neural network for classification and prediction

Comparing with conventional land use modelling tools, ML techniques excel at processing complex variables and predicting dynamic patterns. At the first stage, a Convolutional Neural Network (CNN) model is used to identify past growth patterns of informal settlements among various land use types, which are extracted from satellite images. U-Net, as a type of CNN models, was initially created for biomedical image segmentation (Ronneberger, Fischer, & Brox, 2015). U-Net is constructed by multiple down-sampling and up-sampling layers, which are connected by concatenate layers to enable the outputs have high-resolution features. U-net is known of excelling at producing precise segmentations with fewer training images and less training time. Considering the fact that the data set in this research is manually annotated by the author, it is very time-consuming to form an enormously large data set. Therefore, U-net is chosen for the task of land use classification.

Regarding the ML application for land use prediction, the advantages of implementing artificial neural networks (ANN) in land use prediction have been widely recognized. However, one of the drawbacks of conventional ANN is that it processes values in each cell independently, which means that the spatial relationships among pixels are neglected. This drawback of ANN is resolved in CNN models, which can process matrix and tensor data. Inspired by Batty's (2018) argument, the essence of ML is iterative procedures of searching for a combination of weights and elements that could perfectly reproduce the original combinations, despite that CNN has been mainly applied for image classification rather than prediction, considering its capability of processing multi-dimensional spatial-related data, this research proposes a method of using a CNN model to predict the growth patterns of informal settlements. During this stage, multiple dimensional driving factors of land use changes are inputted into the model as independent features. The coding language selected for building CNN models is Python and the ML packages used are Tensorflow and Keras, which are widely-accepted packages for building ML models efficiently.

3.2. Entropy weight method

Since flood is the most frequent and intense natural hazard that Shenzhen is prone to, this research is focused on mapping the flood susceptibility of the study area. Flood susceptibility is a compound of the influence of multiple contributors. Consequently, a method for synthesising effects of multiple variables is required. The widely-accepted methods for solving multi-criteria tasks include the Analytic Hierarchy Process (AHP), entropy weight method, and Logistic Regression (LR). However, AHP has been criticised of its subjective process that heavily relies on expert knowledge (Shi et al., 2018), and LG requires a target data for training, in this case, hazard inventory data have to be acquired as a training target (Pradhan, 2010). Therefore, the entropy weight method is selected because it is a relatively objective method and can be applied without the input of disaster inventory data. The general process of conducting the entropy weight method is as follows: first, the information entropy value is computed

$$e_i = -\frac{1}{\ln(n)} \sum p_{ij} \ln p_{ij}, \quad p_{ij} = \frac{x_{ij}}{\sum x_{ij}}$$

according to the equation:

(1)

where e_i is the entropy of i^{th} factor, x_{ij} represents the j^{th} vaule of i^{th} factors. n is the number of features is applied here to ensure the value of e_i is between 0 and 1. Then the weight of each factor can be calculated as follows:

$$w_i = \frac{1 - e_i}{\sum (1 - e_i)} \tag{2}$$

Where w_i is the weight of i^{th} factor. Therefore, an overall weighted result can be calculated by the sum of multiplication of the values of each factor with its corresponding weight.

4. Results and discussions

4.1. Informal settlements segmentation and growth prediction

4.1.1. Informal settlements segmentation

A U-net model was built to segment pixels of informal settlements from aerial images, as well as additional four land use types, including formal settlements, vegetation, waterbodies and impervious surfaces. The architecture of U-net built in this research consists of four down-sampling layers and four up-sampling layers, the number of channels in the first convolutional layer was 32, and the kernel size was 3×3 in all of the convolutional layers in the model.

The original data set contains 18 images that each has 4,096 rows with 4,096 columns of pixels, each of the original images was resampled to 3,000 images that each has 256 rows with 256 columns of pixels. Such resampling can enable desktop computers to process the data set. The data set was then divided into a training set and a validation set with a split ratio of 0.33. The settings of the model are as follows: batch 200 and epoch 50,0000; after nearly six hours of training time, the model achieved 94.97% of training accuracy and 91.50% of validation accuracy. Based on the trained model and weights, the spatial distributions of all the informal settlements across the whole city of Shenzhen were identified. Considering the capacity of computer power, an aerial image of Shenzhen with 1,406 km^2 coverage was evenly divided into 16 smaller images to make the actual prediction. It is worth mentioning that, considering the capacity of computer power, although the number of pixels in the results were largely decreased through a max-pooling layer, the features that are essential for the next step analysis has been well preserved. Through this process, land use classification of Shenzhen in the year 2008 and 2018 has been achieved (Fig. 1a-b).

Based on the land use classification of 2008 and 2018, it can be observed that there are no significant changes happened in the patterns of informal settlements in the city centre over this decade, whereas the north-east area shows a considerable amount of growth. Therefore, the next stage of informal settlements prediction is focused on this particular patch, which has a coverage of 351 km^2.

4.1.2. Informal settlements growth prediction

A key assumption at this stage is that the trend land use growth of a certain period of the past would continue for the next same length of a period in the same area. Based on this assumption, the classified land use data of 2008 and 2018 from the previous stage were used as training data to train a CNN predictor, which can project the land use distribution in the year 2028. This CNN predictor consisted of four 2D convolutional layers, and there were four batch normalisation layers after each of the convolutional layers. In terms of parameters of this CNN model, the number of channels in the first convolutional layer was 64, kernel size was 3×3, batch size was 100, and epoch was 500.

In terms of the training data set, the land use data was transferred in to a six-layer one-hot tensor, then was combined with five more spatial features that were regarded as the driving factors of land use changes, including slope gradient, slope aspect, elevation, distance to waterbodies, distance to transport networks. The training accuracy of the model reached 95.15% whereas a validation accuracy of up to 93.84% could be achieved.

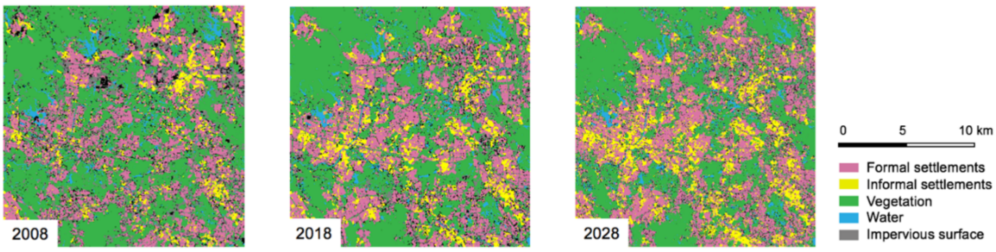

Figure 1. Classification of Land use in a) 2008 and b) 2018, and c) the prediction of land use in 2028

Based on the trained model, the land use of 2018 was combined with the geographic features to predict the land use of 2028, as shown in figure 1c. The predicted land use map of 2028 indicates that informal settlements will continue growing whereby the corresponding area of informal settlements will enlarge to 131% with respect to the year 2018. As for the changes in formal settlements, the growth from 2008 to 2018 is 117%, and the expansion from 2018 to the predicted result of 2028 is 106%. Unlike the growth in urban settlements, vegetated areas experienced a shrinkage from 2008 to 2018, the trend continued with a similar rate from 2018 to the projected year 2028. Interestingly, the predicted patterns of urban growth show of agglomeration in some areas and fragmentation in other areas. Also, although the tendency of land use changes shows consistency in each land use class, the rates of growth and shrinkage were slower.

	Informal settlements	Formal settlements	vegetation	waterbodies	Imperious surface	background
Year 2008	22.74 km^2	86.3 km^2	160.12 km^2	11.86 km^2	44.71 km^2	25.29 km^2
Year 2018	34.01 km^2	100.6 km^2	149.70 km^2	13.56 km^2	32.16 km^2	20.92 km^2
Year 2028	44.54 km^2	106.67km^2	140.32 km^2	13.74 km^2	25.77 km^2	19.96 km^2
Growth rate from 2008 to 2018	150%	117%	93%	114%	72%	83%
Growth rate from 2018 to 2028	131%	106%	94%	101%	80%	95%

Table 1. Number of pixels of each land-use type from 2008 to 2028

4.2. Flood susceptibility mapping

Following equation 1 and 2, the information entropy of causative factors of flooding was calculated. These factors include precipitation, elevation, distance to waterbodies, and land cover that coarsely categorised as impervious surface, vegetation and waterbodies. Among these factors, since the land use distribution is dynamically changing over time, the land cover data was updated for estimating the flood susceptibility for each year, whereas the other three factors remained the same as they are relatively static. As a result, the weight of each factor of the year 2008, 2018 and 2028 was computed. Then the flood susceptibility of the study area was generated by adding weighted beneficial factors and subtracting weighted non-beneficial factors, the diagram of the anticipated flood susceptibility of the year 2028 is shown in Figure 2.

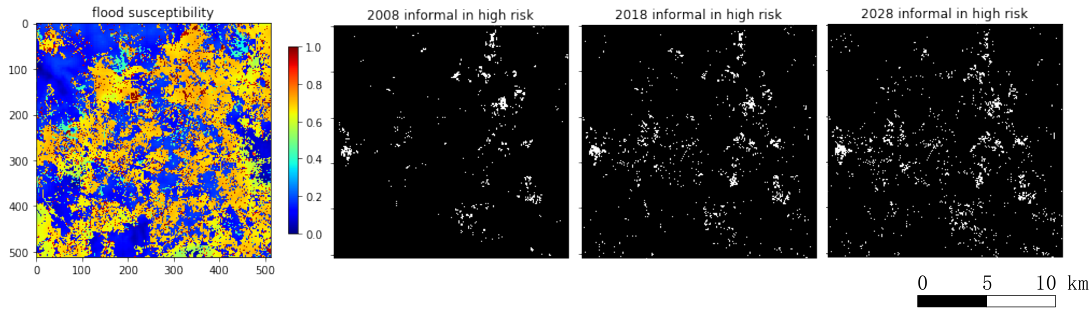

Figure 2. Predicted flood susceptibility and informal settlements which are at high risks in 2028

Based on the flood susceptibility maps from the year 2008 to 2028, the vulnerability of informal settlements could be assessed. As shown in table 2, although the amount of high flood susceptibility area almost remains the same over these two decades, the number of informal settlements is projected to show a substantial growth during this period.

	Area of informal settlements (km^2)	Area of informal settlements inside high flood susceptibility areas (km^2)
Year 2008	22.74	6.05
Year 2018	34.01	9.92
Year 2028	44.54	12.68

Table 2. The number of informal settlements inside high flood susceptibility areas

5. Conclusion and outlook

This research proposed an approach of predicting the growth of informal settlements in the future, as well as a method of estimating the corresponding flood susceptibility for improving the resilience of those emerging settlements. The proposed method consists of three consecutive models, which includes a U-net model for land use classification, a CNN model for land use prediction and an entropy weight method model for mapping flood susceptibility. The one for classifying land use types emphasises the preserving features in high resolution, whereas the other one for the prediction task focuses on the spatial relationships among cells. Both the land use classification and prediction models have achieved highly accurate results. More importantly, the proposed model has showed good performance in predicting potential risk distributions in the study area, as well as indicating the growing risks that emerging informal settlements would need to cope with if the current trend continues.

Regarding the further development, firstly, although the trained prediction model has achieved high accuracy on the data set of the year 2008 and 2018, since there is no ground truth of the land use of 2028, the accuracy of using the trained model to predict future land use cannot be evaluated. To get further insights on model accuracy, future work can include training the model with data of earlier years and use the trained model to predict the land use distribution of an existing year. Secondly, the driving factors of land use changes and causative factors of flooding involved in the modelling process should be further examined about whether all these factors are essential or other factors should be involved. In addition, it also worth discussing that since the morphologies of informal settlements vary from city to city (Taubenböck et al., 2018), the trained model is very unlikely to achieve a highly accurate result for every city across the globe. Although the resampling of training data would be required for such an application for another study area, the method itself can be applied in a generic manner.

References

[1] Kimmelman, M. (2017, April 7). Rising Waters Threaten China's Rising Cities. The New York Times. Retrieved from https://www.nytimes.com/interactive/2017/04/07/world/asia/climate-change-china.html, https://www.nytimes.com/interactive/2017/04/07/world/asia/climate-change-china.html

[2] Pradhan, B. (2010). Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. Journal of Spatial Hydrology, 9(2). Retrieved from http://www.hydromap.com/josh/index.php/JOSH/article/view/81

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv:1505.04597 [Cs]. Retrieved from http://arxiv.org/abs/1505.04597

[4] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

[5] Scovronick, N., Lloyd, S. J., & Kovats, R. S. (2015). Climate and health in informal urban settlements. Environment and Urbanization, 27(2), 657–678. doi: 10.1177/0956247815596502

[6] Shi, Y., Zhai, G., Zhou, S., Lu, Y., Chen, W., & Liu, H. (2018). How Can Cities Adapt to a Multi-Disaster Environment? Empirical Research in Guangzhou (China). International Journal of Environmental Research and Public Health, 15(11), 2453. doi: 10.3390/ijerph15112453

[7] Taubenböck, H., Kraff, N. J., & Wurm, M. (2018). The morphology of the Arrival City - A global categorization based on literature surveys and remotely sensed data. Applied Geography, 92, 150–167.

[8] Yang, L., Scheffran, J., Qin, H., & You, Q. (2015). Climate-related flood risks and urban responses in the Pearl River Delta, China. Regional Environmental Change, 15(2), 379–391. doi: 10.1007/s10113-014-0651-7